Exam 4

April 12, 2016

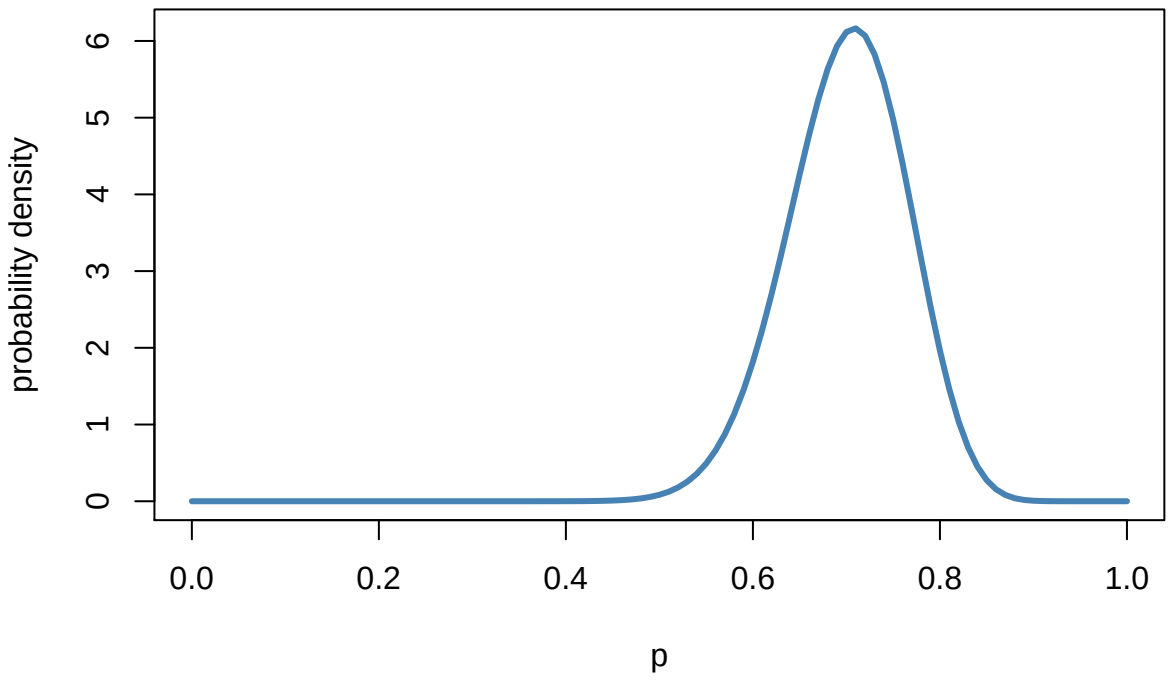
Create a markdown file in RStudio, and call it `YourLastName_Exam4`. Answer all of the exam questions in this file, with your r code for each problem inserted in a separate chunk. When you are finished, knit the file to an .html document `YourLastName_Exam4.html`. Be sure sure your `.Rmd` file compiles properly, and remember to turn in all of your work before you leave. Turn in the following 3 files:

- `YourLastName_Exam4.Rmd`
- `YourLastName_Exam4.html`
- `YourLastName_Data.csv`

If you are taking the exam in class, bring your files up on a flash drive to transfer to my computer. If you are taking the exam at the ACCESS center, e-mail your files to me at ngotelli@uvm.edu.

1a. Using the beta distribution, create and plot a probability density function with a sharp peak at p=0.70. **(5 points)**

```
xVal <- seq(0,1,by=0.01)
beta_pdf <- dbeta(x=xVal,shape1=35,shape2=15)
plot(x=xVal,y=beta_pdf,
     type="l",
     col="steelblue",
     lwd=3,
     xlab="p",
     ylab="probability density")
```



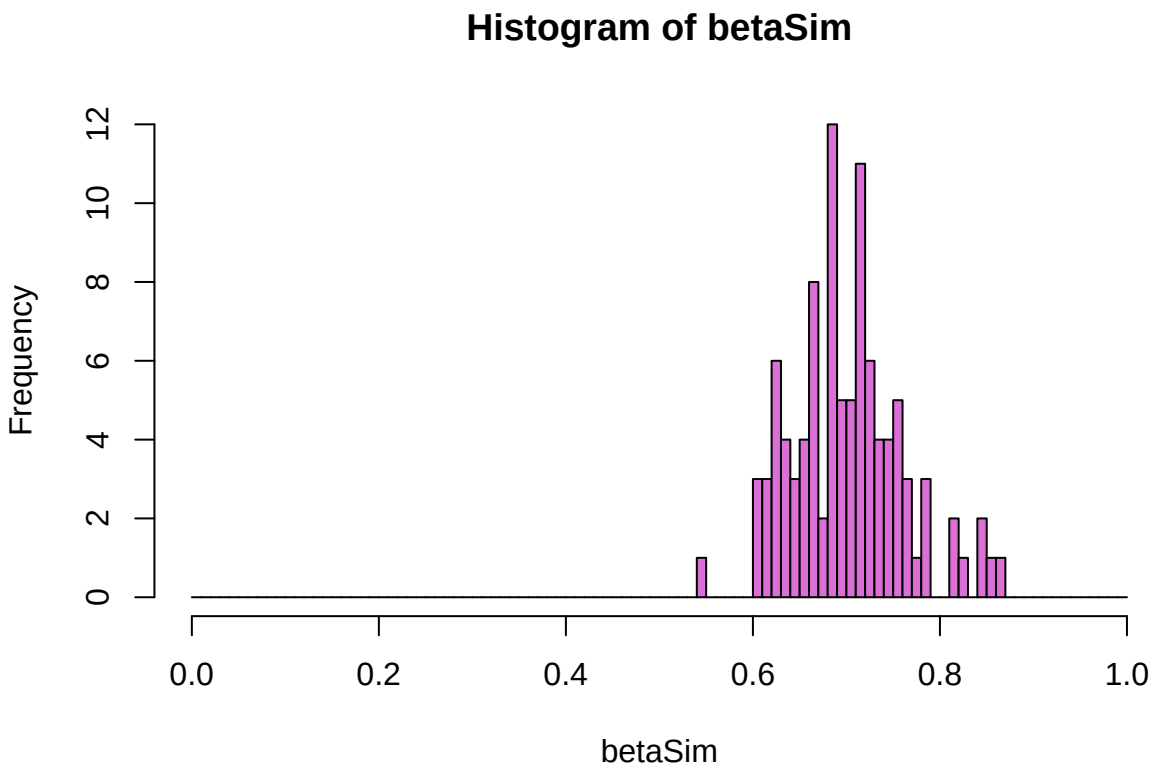
1b. Calculate the exact 95% confidence interval for this distribution. **(5 points)**

```
qbeta(p=c(0.025,0.975),shape1=35,shape2=15)
```

```
## [1] 0.5673703 0.8174806
```

1c. Simulate 100 observations sampled from this beta distribution and illustrate them in histogram plot. **(5 plots)**

```
betaSim <- rbeta(n=100,shape1=35,shape2=15)
hist(betaSim,breaks=seq(0,1,by=0.01),col="orchid")
```



1d. Calculate the 95% confidence interval for these 100 observations. How does this interval compare to the exact interval you calculated in (1b)? **(5 points)**

```
quantile(betaSim,probs=c(0.025,0.975))
```

```
##      2.5%     97.5%
## 0.6024772 0.8461502
```

> The two intervals are similar,but not identical. In this case, the interval estimated from the simulated data is a little narrower than the exact interval from the probability density function.

2a. Using the normal distribution, simulate data for 3 groups with a sample size of 5 observations in each group. The 3 groups should have different means, but the same standard deviation. Bind the data into a single vector. **(5 points)**

```
lowData <- rnorm(n=5,mean=10,sd=5)
medData <- rnorm(n=5,mean=15,sd=5)
highData <- rnorm(n=5,mean=15,sd=5)

dataVec <- c(lowData,medData,highData)
```

2b. Create a vector of a factors with levels ("low", "medium", and "high"). Combine the data vector and the factor vector into a data frame. Print the first 6 lines of the data frame to confirm that it's structure is correct. **(5 points)**

```
myFacs <- factor(rep(c("low","medium","high"),each=5))
myData <- data.frame(dataVec,myFacs)
head(myData)
```

```
##     dataVec myFacs
## 1  8.735172    low
## 2  9.182510    low
## 3 10.948786    low
## 4  9.790304    low
## 5  4.749029    low
## 6 13.320487 medium
```

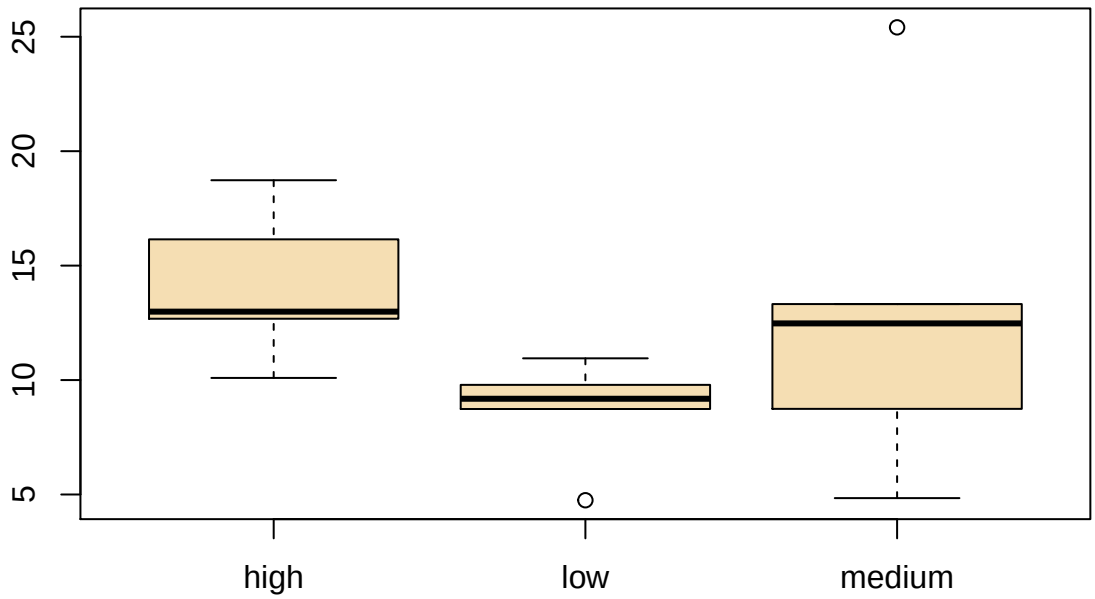
2c. Run an analysis of variance on the data in the data frame, and print out the summary of the results. If necessary, adjust the means and standard deviations of the data groups you created in (2a) so that the ANOVA is ***not*** significant ($p > 0.05$). **(5 points)**

```
myModel <- aov(dataVec~myFacs,data=myData)
print(summary(myModel))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## myFacs       2   82.2   41.10   1.611   0.24
## Residuals   12  306.1   25.51
```

2d. Create a box plot to show the pattern in the data. **(5 points)**

```
boxplot(dataVec~myFacs,data=myData,col="wheat")
```



3a. Using RStudio (create a new Text File, but save as `YourLastName_Data.csv`) or Excel (create a new Excel worksheet and save as a `.csv` file format with the name `YourLastName_Data.csv`), create a data file with two columns, "varX" and "varY". Fill each column with 6 integers. Add a touch of whimsical metadata annotation to the file. Make sure you save the data file in the same location as your `.Rmd` file. **(5 points)**

3b. Read the data into R from the `.csv` file, and analyze it with a simple linear regression model. Print the summary of the regression model. **(5 points)**

```
FakeData <- read.table("Gotelli_Data.csv",
           header=TRUE,
           sep=",")

summary(lm(varY~varX,data=FakeData))
```

```
##
## Call:
## lm(formula = varY ~ varX, data = FakeData)
##
## Residuals:
##        1        2        3        4        5        6
##  0.58344  0.06727  5.24450 -0.41656  1.51876 -6.99741
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6585     3.0258   0.218    0.838
## varX          0.7581     0.3923   1.933    0.125
##
## Residual standard error: 4.452 on 4 degrees of freedom
## Multiple R-squared:  0.4829, Adjusted R-squared:  0.3536
## F-statistic: 3.735 on 1 and 4 DF,  p-value: 0.1254
```